

**Supporting Information for Hargreaves & Eckert *Ecology Letters*:
Local adaptation primes cold-edge populations for range expansion but not
warming-induced range shifts**

1. Additional Methods & Supporting Results

1.1 Study species and sitespg2
Fig. S1. Photos of *Rhinanthus minor* plants and study sites in Kananaskis, Alberta
Table S1. Site and population characteristics
Table S2a. Detailed transplant design and sample sizes for lifetime fitness
Table S2b. Sample sizes for post-emergence success across the high range limit

1.2 Assessing Climate at Plant-height.....pg6
Table S3. Growing season minimum temperatures
Fig. S2. Climate variation among transplant sites 2011-2013

1.3 Between-year Climate Comparisons Above R. minor's Range.....pg9
Table S4. Annual variation in heat accumulation at the highest sites 2011-2013
Fig. S3. Temperatures during our study compared to longer term averages

1.4 Warming Experiment Methods.....pg11
Fig. S4. Warming effect of open topped chambers

1.5 Matrix Models.....pg13
Fig. S5. Population growth rates across *R. minor's* high-elevation range limit

1.6 Dormancy.....pg15
Q1) Does seed dormancy vary systematically among sources or sites?
Q2) Does seed dormancy affect plant performance?

1.7 Additional Performance Results..... pg20
Fig. S6. Performance at HB-mid site in normal vs. truncated growing seasons
Fig. S7. Performance of seeds from the absolute range edge at sites spanning *R. minor's* high-elevation range limit
Fig. S8. Post-emergence performance by year across *R. minor's* high-elevation range limit under natural and artificially warmed conditions
Fig. S9. Plant lifespan
Fig. S10. Minimum growing degree days needed for each source to mature seed
Fig. S11. Effect of warming on fitness components and phenology

2. Statistical Results.....pg25
Table S5. Analyses of performance and phenology of seeds from the range edge vs. the highest large population from two elevational transects (NK & HB)
Table S6. Phenology analyses
Table S7. Analyses of performance across *R. minor's* high-elevation range limit
Table S8. Analyses of performance within the range
Table S9. Analyses of performance and phenology for the warming experiment

3. Additional Discussion.....pg33

1. Additional Methods & Supporting Results

1.1. Study species and sites

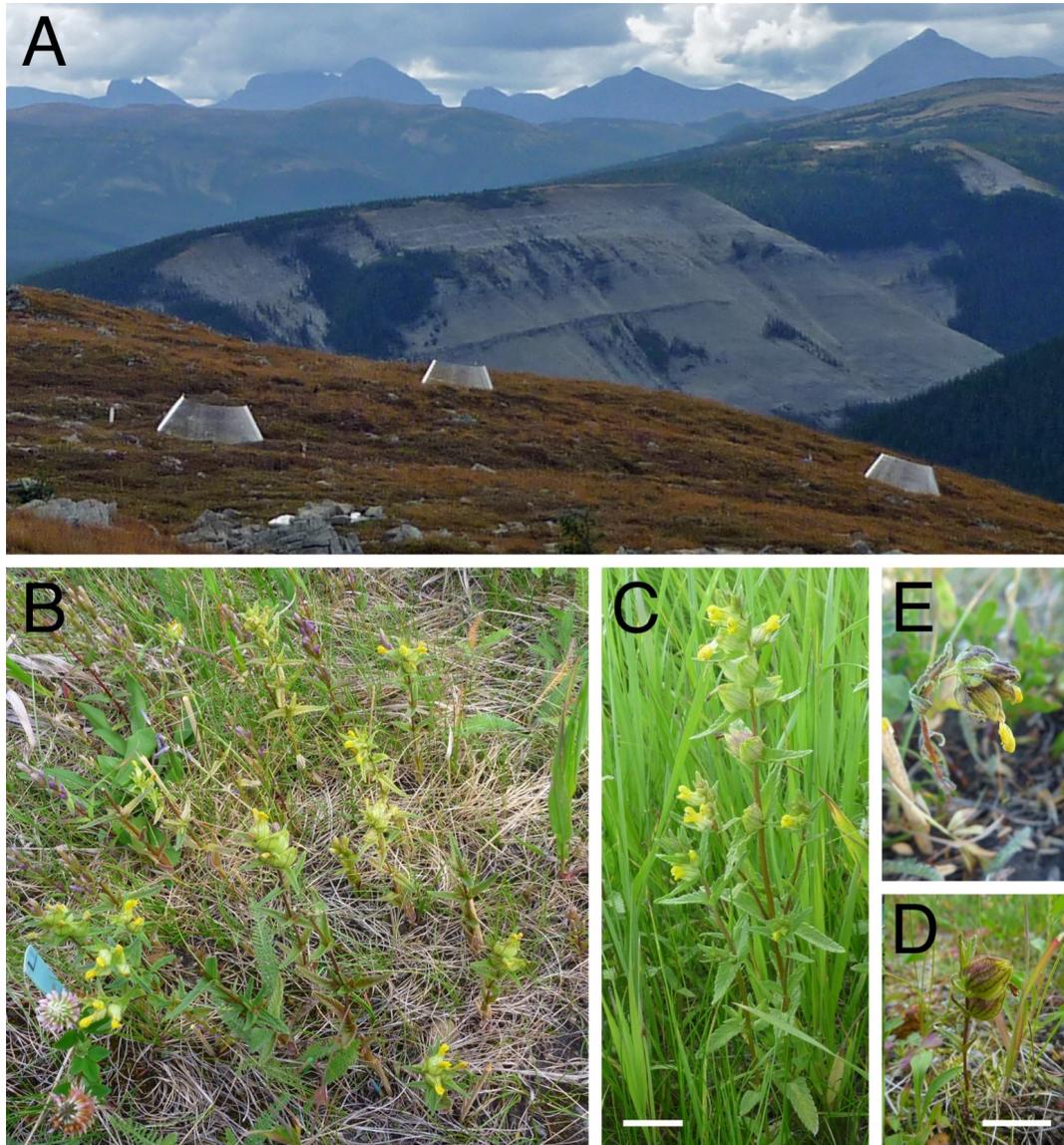


Fig. S1. Photos of *Rhinanthus minor* plants and study sites in Kananaskis, Alberta, Canada. A) Open topped warming chambers above *R. minor*'s range at the HB-transect (*HB-above*, 2380 masl). B) Flowering transplants in 25 x 25 cm subplot of low-elevation seed. C) Low-elevation genotype growing at a low-elevation site, with secondary branches and >20 reproductive structures (buds, flowers, and fruits). D) High-elevation genotype growing at high-elevation site, with only two fruits. Scale bars are 2 cm. E) Late-season flowers killed by frost damage at *R. minor*'s upper range edge.

Table S1. Site and population characteristics. Populations sizes were estimated by assessing plant density along a 4 x 50 m transect through the densest part of the population, and estimating the number of such transects comprising the population. Estimates were done independently by 2 to 3 observers per population, with further transects conducted if necessary to reach a consensus. Size estimates are therefore coarse, but the relative sizes of the populations are reliable.

Transect	Source population – transplant site	Elevation (m)	Habitat type	Clearing date	Latitude	Longitude	Population size
NK	Low – low	1450	Old field	>50 yr	50.941	-115.139	4000
	Mid – mid	1850	Ski hill	1982 ¹	50.950	-115.174	10,000 ²
	High – NA	2130	Ski hill	1982 ¹	50.948	-115.184	4000
	Edge – edge	2225	Ski hill	1982 ¹	50.947	-115.187	1000
	NA – above1	2350	Meadow	Natural	50.946	-115.191	0
	NA – above2	2500	Meadow	Natural	50.945	-115.192	0
HB	Low – low	1350	Roadside	>50 yr	50.269	-114.294	5000
	Mid – mid	1750	Meadow	Unknown	50.214	-114.429	2000
	High – high	2175	Meadow	Natural	50.206	-114.448	2000
	Edge – edge	2320	Meadow	Natural	50.207	-114.458	<30
	NA – above	2380	Meadow	Natural	50.198	-114.458	0
MT	High – NA	2010	Meadow	Natural	50.919	-114.774	2000
FT	High – NA	2000	Roadside	Natural extended late 1960s	50.817	-115.193	500

1) Nakiska ski runs were cleared in 1982 from coniferous forests, but natural *R. minor* populations occur on other aspects of the same mountain up to tree line.

2) Population not discrete, but continues more sparsely up and down the hill.

Table S2a. Detailed transplant design and sample sizes for lifetime fitness. Cell values indicate the sources planted at each transplant site for each of three growing seasons; 2012 and 2013 include additional plots at some sites planted for the experimental warming treatment. Sources abbreviations: low (L), mid (M), and high (H) elevation sources local to the transect they were planted on, the highest population along the transect (i.e. the upper range edge: E), and high-elevation populations from Moose Mt (MT) and Fortress Mt (FT).

Transect	<i>n</i> plots (sources planted in each plot) ¹				
	2011	2012		2013	
Elevation (site)	Control	Control	Warmed	Control	Warmed
NK transect					
1450 (low)	10 (L, M, H)	10 (L, M, H)		5 (L, M, H)	
1825 (mid)	10 (L, M, H)	10 (L, M, H) +10 (M, H) ³		5 (L, M, H)	
2250 (edge)	10 (L, M, H, E, MT, FT)	10 (L, M, H, E, MT, FT)	10 (M, H)	10 (L, M, H, E, MT, FT)	10 (M, H, MT)
2350 (above1)	10 (L, M, H, MT, FT)	10 (L, M, H, MT, FT)		0	
2500 (above2)	10 (L, M, H, MT, FT)	10 (L, M, H, MT, FT)	10 (M, H)	10 (L, M, H, MT, FT), 5 (E)	10 (M, H, MT)
HB transect					
1360 (low)	10 (L, M, H)	15 (L, M, H) ²		1 (L, H), 2(M)	
1750 (mid)	10 (L, M, H)	10 (L, M, H) +10 (M, H) ^{3,4}		5 (L, M, H)	
2175 (high)	10 (L, M, H, MT, FT)	10 (L, M, H, MT, FT) +7 (M, H) ⁴	3 (M, H) ³	10 (L, M, H, MT, FT)	10 (M, H, MT)
2320 (edge)	10 (L, M, H, E, MT, FT)	0		0	
2380 (above)	10 (L, M, H, MT, FT)	10 (L, M, H, MT, FT)	7 (M, H)	10 (L, M, H, MT, FT) +5 (M, H, MT) +4 (M, H) ⁴	10 (M, H, MT) +1 (M, H) ⁴

1) All plots, including those in the warming treatment, were included in analyses of emergence.

2) Five extra plots were planted as insurance against cattle trampling. One Low subplot was deemed too contaminated by natural seed and excluded from all analyses.

3) For 2012 we originally planted 10 plots intended for the warming experiment at each mid-elevation site, as well as the range edge and above-range sites. At NK we set up OTCs but abandoned the experiment due to setup problems. These 10 plots are included only

in emergence analyses. At HB-mid all 10 plots planted for the 2012 warming treatment were instead incorporated into the control treatment (i.e. OTCs were never deployed), and are included in all performance analyses.

4) Due to poor emergence of HB-High seeds in 2012: a) All 10 plots planted for the warming treatment at HB-mid were incorporated into the control treatment in 2012. b) Seven of the 10 M&H plots planted for warming at HB-high in 2012 were incorporated into the control treatment. c) At HB-above we planted 20 potential open-topped chambers (OTC) warming plots for 2013 to ensure having 10 plots with seedlings for the warming treatment. OTCs were put on the plots with highest emergence. Survival from emergence to reproduction did not correlate with emergence at HB-high and HB-above (binomial GLMM on original control plots, full model $\text{prop_reproduction} \sim \text{prop_emergence} + \text{Site-Elevation} \times \text{Source} \times \text{Year}$, likelihood-ratio test for the effect of prop_emergence : $\chi^2_{df=1} = 0.95, P = 0.33$), so transferred plots were retained in analyses of proportional reproduction and seeds/reproductive plant, but not lifetime fitness

Table S2b. Sample sizes for post-emergence success across the high range limit (i.e. the number of subplots per source with at least one emerged plant). Sources abbreviations: low (L), mid (M), and high (H) elevation sources local to the transect they were planted on, and high-elevation populations from Moose Mt (MT) and Fortress Mt (FT) (edge sources not shown for simplification).

Transect	<i>n</i> subplots per source with at least 1 emerged plant				
	2011	2012		2013	
site	Control	Control	Warmed	Control	Warmed
NK transect					
edge	10 all sources	10 all sources	10M, 10H	10L, 10M, 9H, 10MT, 10FT	10M, 10H, 10MT
above1	9 all sources	10L, 8M, 10H, 8MT, 8FT		0	
above2	10L, 8M, 10H, 9MT, 9FT	9L, 9M, 10H, 9MT, 8FT	10M, 9H	5L, 7M, 8H, 8MT, 9FT	4M, 5H, 9MT
HB transect					
high	9L, 8M, 9H, 9MT, 8FT	9L, 11M ¹ , 9H, 9MT, 8FT	3M, 3H	10 all sources	10M, 10H, 10MT
edge	10L 9M 6H 10MT 10FT				
above	6L 4M 7H 5MT 6FT	6L, 4M, 2H, 3MT, 4FT	5M, 2H	5L, 12M ¹ , 6H, 9MT, 6FT	8M, 8H, 9MT

1) In 2012 and 2013 *n* can exceed 10 along the HB-transect as: a) 7 of the 10 M&H plots planted for the warming treatment at HB-high in 2012 were incorporated into the control treatment, b) we planted 20 potential open-topped chambers (OTC) warming plots for 2013 to ensure having 10 plots with seedlings for the warming treatment. Extra plots that did not get OTCs were retained in analyses of post-emergence performance.

1.2. Assessing Climate at Plant-height.

Field Methods: At each transplant site we installed one or two HOBO Pro v2 2-channel External Temperature Loggers (Onset Computer Corp.) in fall 2010. HOBO sensors were suspended in white PVC tubing (12 cm long, 2.5 cm diameter) to shield them from direct sun, and staked to the ground. HOBOS suffered from wire destruction by mammals and occasional mechanical failure, so data coverage varies among sites. In 2012 iButton temperature sensors (DS1921G; Maxim Integrated) were installed at all sites. To buffer them from direct solar radiation and heat transfer from the ground, iButtons were surrounded with 2 mm of Styrofoam then secured with white electrical tape to 20 cm nails. Nails were pushed upright into the ground so that iButtons were 2 cm above the ground, facing north. iButtons were placed beside HOBO temperature sensors to calculate the difference in iButton vs. HOBO measurements, with additional iButtons placed in some sites. HOBOS were set to record every hour, whereas iButtons were set to record every hour during the growing season and every 4 h during the winter to preserve memory.

Data manipulation: While equipment use is consistent among sites, HOBO and iButton coverage varied among years as new equipment was added. When both were available we used their average. HOBOS tended to measure higher maximum temperatures. To improve comparability among years, for each site we calculated the ratio of iButton:HOBO measurements when both were available (i.e. an iButton had been placed beside a working HOBO probe), and used this to standardize iButton-only measurements from that site.

From HOBO and iButton data and field observations of snow melt, we estimated the growing season start date, end date, and duration (days) at each site. *Rhinanthus minor* plants were unaffected by burial in snow for up to 2 d and did not show consistent damage after light frosts. We therefore considered the growing season to be bounded by either a hard frost (≤ -4 °C lasting ≥ 3 d) or an insulating snowpack (indicated by a constant temperature between -2 and 2 °C for 72 h). Hard frosts killed leaf, flower and calyx tissue, but did not always harm developing seeds if they were close to maturation. The estimated growing season is therefore that of photosynthetic tissue and flowers, and seed maturation may extend beyond this.

We calculated the number of growing degree days (GDD) for each day at each site as the average of the daily maximum (T_{\max}) and minimum (T_{\min}) temperatures minus a 'base temperature' ($T_{\text{base}}=10$ °C) below which plants are assumed not to grow, i.e.: $(T_{\max}+T_{\min})/2 - T_{\text{base}}$. (Womach 2005). Any temperatures below T_{base} are set to T_{base} , and T_{\max} is capped at 30 °C, because plants do not generally grow faster at temperatures >30 °C (Womach 2005). Thus a maximum of 20 GDD is possible in one day (if $T_{\min}=T_{\max}$: $(30+30)/2-10=20$), and days that never warm up to T_{base} do not contribute any GDD. The choice of $T_{\text{base}}=10$ °C is common but in this case arbitrary, as we had no *a priori* knowledge of whether *R. minor* develops at lower temperatures. We use GDD as a heuristic to assess relative heat accumulation patterns and needs among sites and sources, which are unlikely to change with choice of T_{base} . However, even with a correct T_{base} , this method can overestimate daily heat sums when minimum temperatures are below T_{base} as more hours will be spent below the growth threshold (Cesaraccio *et al.* 2001). $T_{\min} < T_{\text{base}}$ more often and T_{\min} was lower during growing seasons toward higher elevations (Table S3), so true differences in GDD among sites may be even more extreme than we detect.

For each growing season we determined the total GDD, and mean July temperature. GDD were summed within the estimated growing season (i.e. the frost and snow-free period

defined above). We calculated mean July temperature because all sites have developing *R. minor* this month, whereas low sites have finished by August and high sites are usually still snow covered in June. We have growing-season climate data for three years (2011-2013) at most sites, except 2012 at NK-mid and NK-above1 and 2013 at NK-low.

For each winter we calculated the absolute minimum temperature close to ground level, which could potentially damage seeds through freezing, and the total days of snowpack. Snowpack protects the ground from desiccation and its melt contributes greatly to spring soil moisture, so sites with greater snow pack may have higher winter and spring soil moisture, potentially enhancing emergence. At most sites snowpack was defined as days in which the temperature remained between 2 and -5 °C, and did not vary by >5 °C, indicating snowpack was deep enough to insulate the ground (this temperature range is slightly wider than that used to define the end of the growing season to allow for additional cooling during prolonged cold spells). The two ski-hill sites, NK-mid and NK-edge, were clearly covered by thick snow for the entire winter, but the snowpack lost its insulating properties in mid-winter due to compaction. For these sites we included mid-winter days with a wider temperature range. We have climate data from four winters (2010-2014) for most sites, but are missing winter 2011-12 for HB-above and 2013-14 for NK-above2.

Analyses: Models considered a single climate variable as the response, and siteID as the explanatory factor; both transects were included in models so sites could be directly compared (full model: `climate ~ site`). Models of count data (days in growing season, days of snowpack, and GDD in the growing season; GDD behave like count data as they are bounded at 0) used Poisson distributions (`glm` function, base R). Temperature data (minimum winter temperature and mean July temperature) were analysed using linear models (`lm` function). Significance of siteID was determined using likelihood ratio chi-squared tests comparing the full model to that with just an intercept. When siteID was significant we used least squared means (`lsmeans` function, `lsmeans` package 2.25-5 (Lenth 2016)) to test for pairwise differences among the 10 sites, with the Tukey HSD test to maintain an overall α of 0.05. Site differences in mean plant-height growing season climate, determined from HOBO and iButton data, are in Fig. S2.

Table S3. Growing season minimum temperatures. The $(T_{\max} + T_{\min})/2 - T_{\text{base}}$ formula may overestimate GDD when $T_{\min} < T_{\text{base}}$. Using the daily T_{\min} data during July and August 2011-2013 (i.e. when all sites were active), we tested whether sites differed 1) in how often $T_{\min} < T_{\text{base}}$ (binomial GLMM, ‘yes’ if $T_{\min} < T_{\text{base}}$, `~ Site + (1|year) + (1|sensor.type)`), and 2) in T_{\min} itself (GLM: `~ Site + (1|year) + (1|sensor.type)`). Significance determined as above. Sites that do not share a letter are significantly different.

Response	Effect of Site	Significant contrasts (sites ordered from warmest to coldest)
1) Frequency $T_{\min} < T_{\text{base}}$	$\chi^2_{\text{df}=9}$: 1105, $P < 0.0001$	HBL < NKMb HBHbc NKLbc HBMc HBAcd NKEcd HBE d=NKA1d=NKA2d
2) T_{\min}	$\chi^2_{\text{df}=9}$: 553, $P < 0.0001$	HBL > NKM=HBH=HBM=NKL > NKE=HBA=HBE > NKA1=NKA2

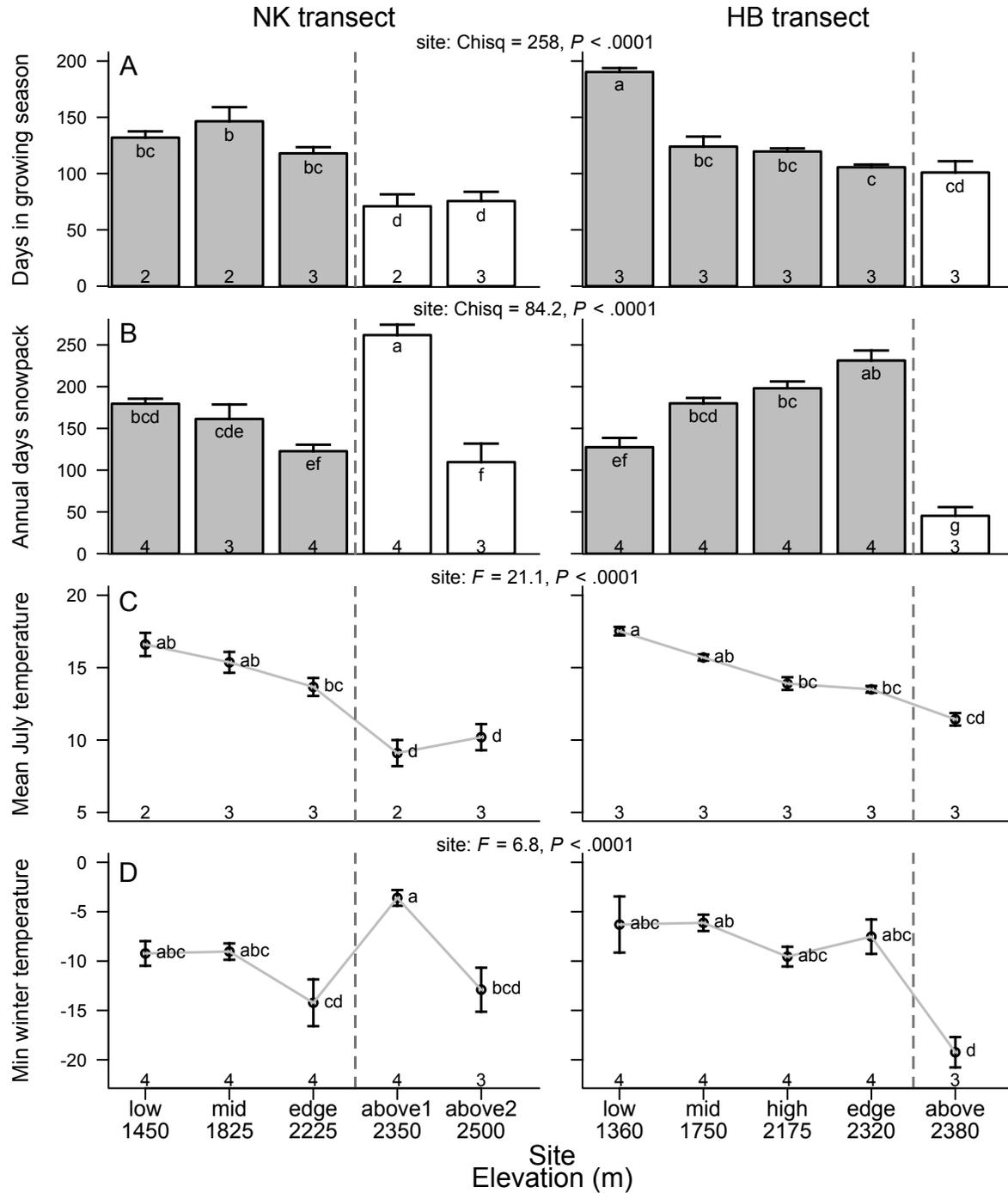


Fig. S2. Climate variation among transplant sites 2011-2013 (mean \pm SE): A) growing season length, B) days of semi- to fully-insulating snowpack, C) mean July temperature ($^{\circ}$ C), D) minimum winter temperature ($^{\circ}$ C). Model results shown at top of each row, $df = 9$, contrasting letters indicate significant site differences across transects. Numbers at bottom indicate the sample size of seasons; data usually include three growing seasons (2011–2013) and four winters (2010–2014), exceptions detailed in Climate assessment methods. Sites right of the dashed line (white bars) are above the range

1.3. Between-year Climate Comparisons Above *R. minor*'s Range

To directly compare climate at the highest sites (HB-above and NK-above2) among study years, we used records from provincial and university weather stations on the same peak: the Hailstone Butte fire tower (2330 m, 900 m linear distance from HB-above) and the University of Saskatchewan Centre for Hydrology Centennial ridge weather station (2500 m, at the top of NK-above2). We used these rather than federal weather stations as they had better data coverage in 2012 and 2013 at NK-above2 and no federal station was available for HB-above, but did not use them for long-term comparisons as records do not cover as many years. Weather station data at HB-above were only available for dates when the fire tower was staffed. As these dates differ among years, we limited calculations to GDD / growing season rather than monthly averages.

To assess whether climate during our study were typical of longer-term climate averages, we used mean daily temperature records from the nearest federal weather station (Environment and Climate Change Canada) that was within 25 km of study sites and had records up to 2013. Appropriate stations were available close to NK-above2, NK-low, and HB-low. We used data already compiled into monthly averages when available, otherwise we calculated monthly averages from daily means including only months with records for ≥ 20 d. For each month (Jan-Dec) we then calculated the mean and SD across years (earliest available year to 2010). For study years (2011–2013), we calculated monthly averages from daily means including only months with records for ≥ 15 d (a slightly lower threshold than for non-study years). Results are below (Fig. S3).

Table S4. Annual variation in heat accumulation at the highest sites 2011-2013. Growing degree days (GDD) / growing season are calculated from weather station data for air temperature > 1 m from ground, so not directly comparable to temperatures measured at plant height, and not necessarily comparable between sites given different collection procedures (e.g. post emergence performance of MT and FT seeds planted at both transects does not suggest that HB-above was always warmer than NK-above2; Fig. S2).

Site (elevation)	Total GDD in growing season		
	2011	2012	2013
NK-above2 (2500 m)	116.3	181.9	193.1
HB-above (2380 m)	202.2	263.5	279.5

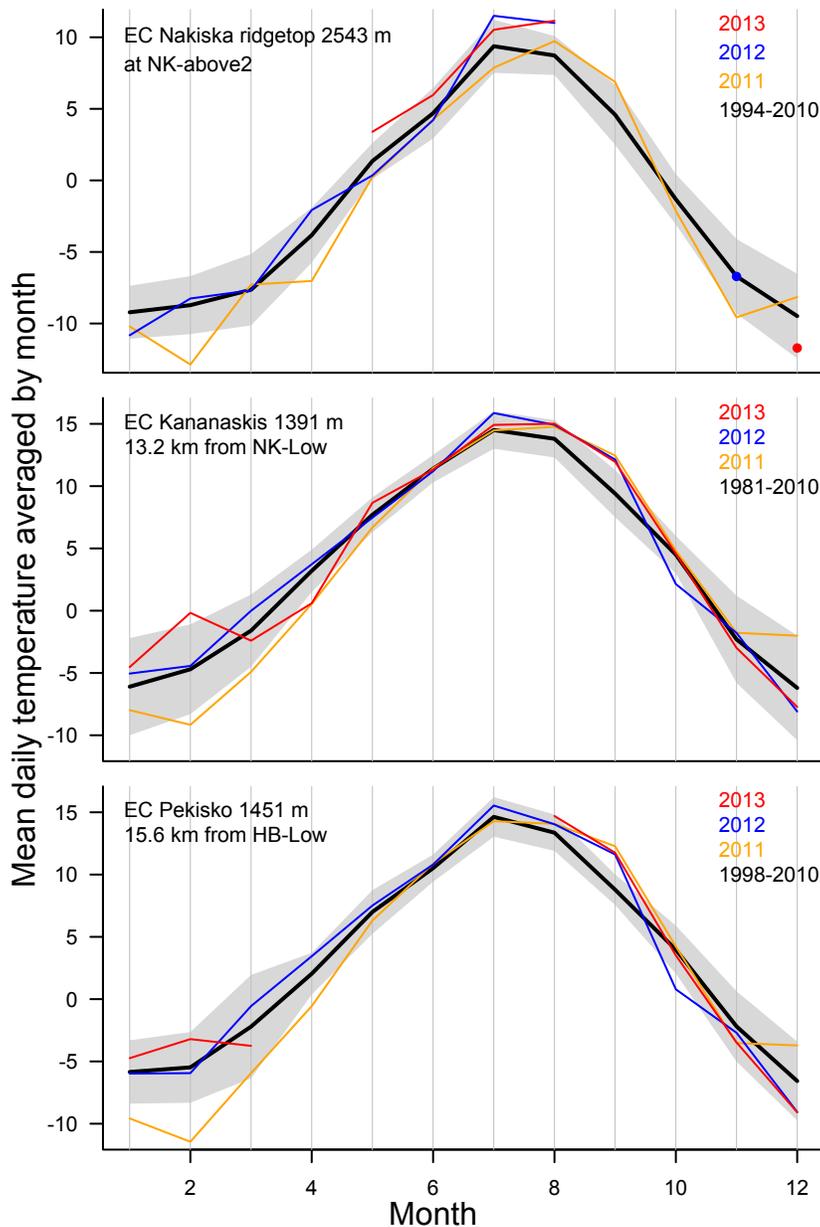


Fig. S3. Temperatures during our study compared to longer term averages, from long-term weather stations close to transplant sites. Growing season temperatures recorded during study years (coloured lines/points) are within 1 SD (grey shading) of long-term means (black lines), but were warmer than average during mid-summer 2012 and 2013 at NK-above2 and in September at all sites and years with data. Growing season data only includes months with records for ≥ 20 d, resulting in gaps for some years. Top left text indicates the Environment and Climate Change Canada station name, elevation, and proximity to the nearest study site. Nakiska ridgetop station was also the closest to NK-low, but the Kananaskis station is at a more comparable elevation.

1.4. Warming Experiment Methods

OTC construction: Open-topped warming chambers (OTCs) were constructed from 1.2 x 2.4 m sheets of 1.5 mm thick Lexan plastic, cut to form a cone with a ~1.2 m diameter base, a 1 m diameter opening, ca. 0.4 m tall. We chose this relatively low, wide, and steep-sided design as these features minimize unwanted effects of OTCs (extreme temperatures, altered precipitation and soil moisture), although they also reduce the warming achieved (Marion *et al.* 1997). We also ensured that plots were positioned in the center of OTCs and so received direct sunlight and vertical precipitation. As *R. minor* can set a full complement of seeds via autonomous self-pollination (Hargreaves *et al.* 2015), effects on pollinators were not a concern.

Assessing OTC effects on temperature: To measure OTC's warming effect we placed iButtons in the center of 5 to 13 OTCs and an equal number of control plots each year (2012, 2013) at all sites in the warming experiment (NK-edge, NK-above2, HB-high, HB-above). iButtons were usually deployed the day OTCs were set up, but sometimes the following week. iButtons were set up as in 'Climate assessment methods' and recorded every hour. From these data we calculated GDD for each day as above.

For each OTC site we compared daily maximum, minimum, and mean temperature and daily-GDD between OTC and control plots (Fig. S4B). We used one linear mixed effects model per site, with Warming treatment, Year, their interaction and Julian date as fixed effects and plotID as a random intercept. To account for temporal models included an autocorrelation error structure of order 1 with a continuous time covariate (Zuur *et al.* 2009) using the nlme package version 3.1-137). Full model: `response ~ Warming*Year + Jday, random=~1|plot, correlation= corCAR1(form=~Jday)`, where response is daily Tmax, Tmin, Tmean, or GDD. Significance of Warming treatment was determined for each year using lsmeans (`lsmeans(model, pairwise ~ Warming|Year)`).

In 2012 we conducted additional 'point' temperature measurements in warmed and control plots at our two highest sites (NK-above2 and HB-above; Fig. S4A). On one afternoon per site (July 29 and Aug 1, both mixed sun and cloud) we measured soil and air temperatures at each subplot of High or Mid seeds: soil temperature 2 cm below the ground surface, measured with a hand-held digital thermometer probe; air temperature at ground level (1 cm above soil surface) and 10 cm above ground surface, measured with a hand held thermometer. As air temperature changed over the course of the afternoon we alternated between control and warmed plots to ensure this did not introduce bias between the treatments. Data were normally distributed and analysed together using one linear model per temperature variable with site (NK-above2 or HB-above), treatment (warmed or control), and their interaction as fixed effects.

The warming effect of OTCs was clearly visible in both the point and continuous temperature measurements (Fig. S4). At the highest sites both soil and air temperatures were warmer in OTCs, though the difference was not significant for soil temperature at NK-above2 (Fig. S4A). As found by other studies (Marion *et al.* 1997), OTCs produced the strongest warming in the day (larger and more often significant difference in T_{\max} than T_{\min} ; Fig. S4B). While OTCs have occasionally been found to reduce nighttime temperatures (Marion *et al.* 1997), T_{\min} was always warmer in our OTCs vs control plots, with the exception of 2012 at HB-above when mean minimum temperatures did not differ between treatments (Fig. S4B). Accordingly, plants in OTCs experienced more GDD at each site in each year, though the difference was not significant at HB-high in 2013.

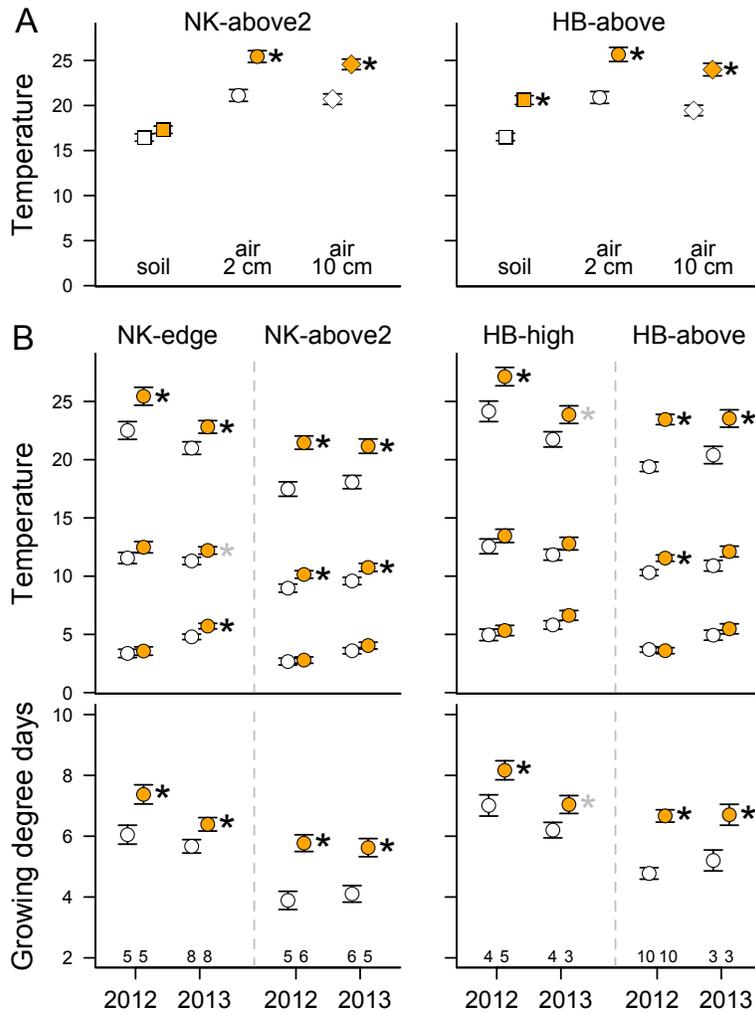
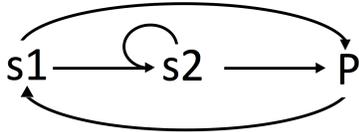


Fig. S4. Warming effect of open topped chambers (OTCs). White points are unmanipulated control plots, orange points are plots in OTCs, sites right of the dashed line are above *R. minor*'s range. Data show least squared means \pm SE extracted from models; * indicates significant difference between warming treatments (grey * indicate treatment significance was between $P=0.05$ and 0.06). (A) Instantaneous temperature of soil 2 cm below surface, air 2 cm above ground (as per iButtons in B), and air 10 cm above ground, measured once in each Mid- and High-seed subplot on one afternoon per site in 2012 (n Control/OTC: NK = 20/20, HB = 20/14, OTC sample size reduced as OTCs had been removed from plots where no plants remained). Linear model: temperature \sim Warming \times Site. (B) Daily maximum, mean, and minimum temperature (upper, middle, and lower point pairs, respectively) and daily GDDs calculated from max and min temperatures. Measurements are from 2 cm above ground by iButtons in the plot centre. Bottom numbers indicate sample size (#plots with an iButton / treatment / year / site). Note that temperatures are not directly comparable across years as iButtons were deployed later in the 2012 season. One GLMM per response per site: response \sim Warming \times Year + Jday, random= \sim 1|plot, correlation= corCAR1(form= \sim Jday)).

1.5. Matrix Models

To account for seed dormancy in assessing habitat suitability at and above the upper range edge, we determined the instantaneous, density-independent population growth rate (λ) for each source population at the three highest sites along each transect: NK-edge, NK-above1, and NK-above2, and HB-high, HB-edge, and HB-above. Performance was divided into 3 life history stages: 1st year seeds (s1), dormant seeds (s2), and plants (P). We generated a transition matrix according to the diagram below:



1st year seeds can become plants with the probability s1P or dormant seeds with the probability s1s2, dormant seeds can remain dormant with the probability s2s2 or emerge to become plants with the probability s2P, and plants become 1st year seeds with the probability Ps1. We estimated these parameters from the data as follows.

s1P = mean proportion emergence for a given source at a given site across years

s1s2 = mean proportion of seeds that emerged in their second summer, averaged across the three highest sites along each transect, and across sources and years as we did not detect a significant site or source effect on dormancy. Dormancy (seeds that emerged in their second summer divided by the seeds remaining, i.e. that had not germinated in their 1st summer) was higher at NK (7.3%) than HB (3.2%) – see ‘Dormancy’ below.

s2s2 = s1s2 (i.e. assuming dormancy does not change with seed age)

s2P = s1P (i.e. assuming emergence does not change with seed age)

Ps1 = mean seeds produced per emerged plant for a given source at a given site across years

We reran matrix models allowing the dormancy fraction (s1s2 and s2s2) to vary by site and source, but this did not qualitatively change results.

We used a simple simulation to model population growth (R (R Development Core Team 2015)). Populations were initiated with 100 first year seeds, 0 dormant seeds and 0 plants. This vector of states was multiplied to the transition matrix to generate a new vector of states in the second generation, and the total populations size, including seeds and plants, was divided by the population size in the time step before to calculate the growth rate. This was repeated until the growth rate stabilized on a single value, usually 100 generations. Simulations were run longer if growth rates took longer to stabilize, or initiated with fewer or more individuals if populations reached 0 or infinity before growth rates stabilized.

We calculated λ for all sites and sources in the control treatment to assess whether habitat was suitable for each source (Fig. S5). To assess whether warming the growing season was enough to make above-range sites suitable, we also calculated λ for sites and sources in the warming treatment (Fig. S5), using the overall emergence fraction for each site x source combination. Finally, to assess whether warming would make above-range sites suitable if winter conditions were also more like those at the range edge, we re-ran models using Ps1 from the OTC treatment above the range and s1P from the range edge of that transect (Fig. S5).

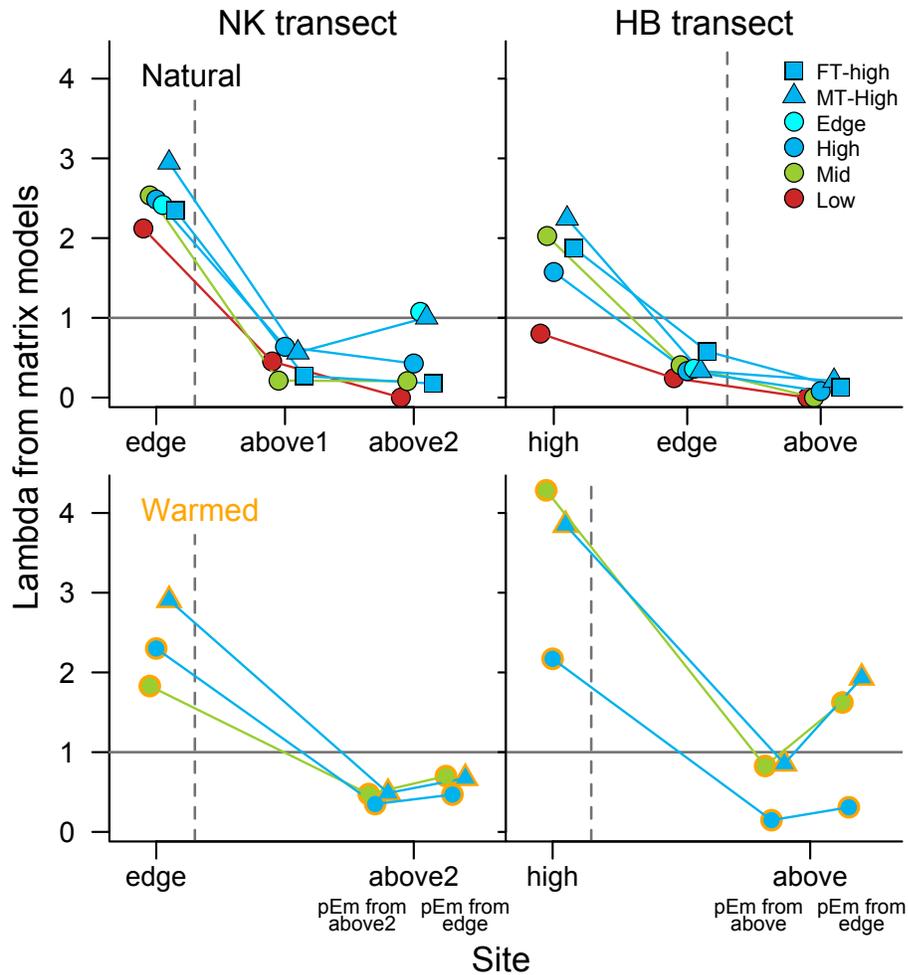


Fig. S5. Population growth rates across *R. minor*'s high-elevation range limit (vertical dashed line) derived from matrix model simulations. Growth rates ≥ 1 (reference line) indicate self-sustaining populations in the absence of demographic variability. Natural: For unwarmed conditions, the transition matrix used mean emergence and mean seeds/emerged plant for each site x source combination across years (2011–2013 except for NK-above1 (2011–2012) and HB-edge (2011 only)), and the overall dormancy rate per transect calculated as the mean emergence in the second summer across sites, sources and years (detailed methods in ‘Dormancy, below’). Low, Mid, High and Edge sources are local to the transect on which they are planted. Note that NK-Edge seeds were only transplanted to the NK-above2 site in the year with the highest success at that site (2013), so results may overestimate their success. Warming: λ estimates under warming use seeds/emerged plant from plots in warming chambers (2012 & 2013). At edge and high sites emergence is taken from 2011–2013 means at those sites (i.e. the values used in top panels). At above-range sites emergence is taken either from the same sites, to assess whether warming the growing season alone would make habitat suitable (left), or from the range edge, to assess whether warming would be make above-range habitat suitable if winter conditions resembled those at the range edge (right). Warming the growing season generally improved performance above the range, but not enough to make $\lambda \geq 1$. Warming plus ameliorated emergence conditions might make the HB-above site suitable for HB-Mid and MT-High seeds.

1.6. Dormancy

In European populations *R. minor* does not form a persistent seed bank, and >90% of seeds germinate after their first winter (Westbury 2004). However, in August 2012 we found plants growing in 2011 plots above the range, which must have arisen from dormant seed as no plants produced seed at these sites in 2011. From this point we monitored past-year plots to quantify Q1) the degree of seed dormancy, and Q2) whether plants from seed that spent at least one summer dormant differed in performance compared to plants from 1st-year seed. Counts of dormant plant emergence from 2012 reflect both post-dormancy emergence and survival as plants were counted late in the season. By the time dormant plants were observed, survival already differed among sources (not shown), introducing potential bias into perceived emergence, so we exclude 2012 data from analyses below. Since HB-edge was only planted in 2011, excluding 2012 data means we have no data on second-summer emergence, the measure least likely to be confounded by accumulating seed death. We therefore also exclude data from HB-edge.

Methods for monitoring and calculating performance of dormant plots were the same as for 1st-year plots. Subplots could not be used for dormancy monitoring/analysis if 1st-year transplants (or, later, plants from dormant seed) had matured and released seed. From field notes we identified subplots in which >2 fruits had opened and eliminated them from analyses for subsequent years. E.g. if plants in subplot X dropped seed in the subplot's second summer, data for year 1 (non-dormant seed) and year 2 (second summer seed) were retained in analyses but data from subsequent years discarded. Dormancy was monitored from 2012–2014 at HB and 2012–2015 at NK, but in 2015 plants were removed as soon as emergence was recorded so cannot be used for analyses of performance.

As for data from 1st year (non-dormant) seeds, dormancy data are analysed separately by transect because the low, mid and high seeds are unique to each transect (i.e. site and source are not fully crossed across transects).

Q1) Does seed dormancy vary systematically among sources or sites?

We analysed whether the proportion of seeds remaining (i.e. the number planted minus any that emerged in previous summers) that produced seedlings each year varied by site, source, the number of years seeds lay dormant, and all possible interactions between these factors. We predicted that: Prediction1) sources from low-quality sites (i.e. high elevations) might have higher dormancy as a fixed bet-hedging strategy (i.e. a significant source effect); Prediction2) all sources might show higher dormancy above the range as a plastic response to low site quality (i.e. a significant site effect); Prediction3) the proportion of seeds remaining that emerge will decline as seed age increases, reflecting increased mortality. We used generalized linear mixed models (GLMMs: glmer command, lme4 package 1.1-13 (Bates *et al.* 2015)) with binomial error distributions and random intercepts for plot (to account for plot effects among related subplots), year (to account for variable emergence conditions in different years), and subplot (to account for overdispersion). The response variable was emergence starting the second year after planting, i.e. yr.emergence ≥ 2 . NK models consider NK-edge, NK-above1 and NK-above2 sites, HB models consider HB-high and HB-above sites. Significance of factors was determined using likelihood ratio tests with and without the factor in question, compared to a Chi-squared distribution. Non-significant interactions were dropped from models. Differences among

significant factor levels were tested for using least squared means (lsmeans function, lsmeans package). Estimates of the effect of continuous factors are from the summary function (base R).
 Full model: `seeds.to.emerge/seeds.remaining ~ Site x Source x Yr.emergence + (1|plotID) + (1|year) + (1|subplotID)`

Nakiska

Sites in model: NK-edge, NK-above1, NK-above2

Sources in model: L, M, H, MT-High, FT-High

Minimum adequate model (MAM):

`Site x Source + Source x Yr.emergence + Site x Yr.emergence + (1|plotID) + (1|year) + (1|subplotID)`

We estimated the global dormancy fraction (i.e. proportion of seeds remaining that emerge each year) using the mean of 2nd-year emergence from remaining seeds, as this year has the lowest build-up of dead and eaten seeds. The estimated dormancy fraction was 0.080 ± 0.0064 , or 8%. If calculated by site and source, estimates vary from 1.4% (MT-High seeds at NK-above1) to 12.5% (FT-High seeds at NK-above2).

Prediction1) Source differences: No source had a consistently different dormancy strategy. The fraction of dormant seeds differed among sources, but the pattern was not consistent among sites and was only significant at NK-above1. Contrast results from `lsmeans(model, pairwise ~ Source | Site):`

At NK-edge: Mid = High = FT-High = Low = MT-High

At NK-above1: Mid = Low = High = FT-High > MT-High (FT>MT $P = 0.06$)

At NK-above2: FT-High = Mid = MT-High = Low = High

Thus we have no evidence for fixed adaptive dormancy strategies among sources.

Prediction2) Site differences: Dormancy did not differ consistently among sites. Each site had the lowest dormancy fraction for at least one source, and dormancy only differed significantly by site for two sources (`lsmeans(model, pairwise ~ Site | Source):`

For FT-High seeds: above2 > above1

For MT-High seeds: above2 = edge > above1

Thus we have no evidence for consistent plastic responses in dormancy as a response to site quality.

Prediction3) Effect of seed age: Dormant seed emergence declined for each source as the number of dormant summers increased, and this decline was significant for most sources. Effect of year of emergence (unbacktransformed estimate \pm SE, $|z$ value|, P , from summary function):

For Low: -0.95 ± 0.18 , 2.65, $P = 0.0080$

For Mid: -0.70 ± 0.16 , 4.57, $P < 0.0001$

For High: -1.03 ± 0.18 , 2.19, $P = 0.029$

For FT-High: -1.42 ± 0.21 , 6.82, $P < 0.0001$

Similarly, the proportion of potentially dormant seeds emerging declined as the number of summers spent dormant increased at all sites. This was most significant above the range (effect of year of emergence, estimate \pm SE, $|z$ value|, P):

NK-edge: -0.87 ± 0.17 , 3.24, $P = 0.0011$

NK-above1: -1.42 ± 0.21 , 6.82, $P < 0.0001$

NK-above2: -0.80 ± 0.14 , 4.45, $P < 0.0001$

Because we did not assess or remove seeds that died or were eaten, the apparent negative effect of years spent dormant on the proportion of remaining seeds to emerge at least partially reflects a buildup of failed seeds in the ‘seeds remaining’ denominator.

Hailstone Butte

Sites in model: HB-high, HB-above

Sources in model: L, M, H, MT-High, FT-High

Minimum adequate model (MAM):

```
seeds.to.emerge/seeds.remaining ~ Site x Yr.emergence x Source +  
  (1|plotID) + (1|year) + (1|subplotID)
```

We estimated the global, second-year dormancy fraction as 0.033 ± 0.0043 - half that of NK, likely reflecting higher seed mortality along HB. If calculated by site and source, estimates vary from 0 (Low seeds at HB-above) to 0.093 (FT-High seeds at HB-High). Excluding Low seeds from the overall mean only raises it to 0.037.

Prediction1) Source differences: The fraction of dormant seeds differed among sources, but only at the HB-above site. Significant contrast results from `lsmeans(lsmeans(model, pairwise ~ Source | site*yrEm))`, showing only sources that differed from at least one other source:

At HB-above: FT-High > MT-High = Low

i.e. FT-High seeds had higher levels of emergence after year 1 than Mt-High and HB-Low seeds. Note that this contrasts results at NK, where dormancy of FT-High seeds was not significantly different from any other source. Thus although we find source differences at HB these do not constitute strong evidence for differing fixed (e.g. genetic) dormancy strategies.

Prediction2) Site differences: In contrast to NK, dormancy differed consistently between sites at HB. Emergence in seeds’ second or later summer was lower at HB-above than HB-high for every source, and this was significant for all sources but FT-High (contrast results from `lsmeans(model, pairwise ~ site | Source*yrEm)`).

Prediction3) Effect of seed age: In contrast to NK, the proportion of potentially dormant seeds emerging did not uniformly decline with seed age across sites and sources (significant Site x Source x Yr.emergence interaction: likelihood ratio $\chi^2_{df=4} = 11.4$, $P = 0.022$). At HB-high dormant seed emergence declined with increasing seed age for three of five sources, but this was never significant. At HB-above, emergence increased with increasing seed age for all sources, but again this increase was not significant for any given source.

Although the proportion of dormant seeds emerging increases with seed age at HB-above, the overall number of dormant seeds emerging is higher at HB-high. This would be consistent with a plastic, bet-hedging strategy, where seeds in poor environments stagger emergence (Donohue *et al.* 2010), although FT and MT seeds, which were planted across the range edge at both transects, showed no evidence bet-hedging at the Nakiska transect. Differential seed mortality at the two sites could explain the overall lower emergence at HB-above, but if seed mortality stayed high this should result in a steeper decline in emergence with year after planting at HB-above vs. HB-high, whereas the opposite was true.

Q2) Does seed dormancy affect plant performance?

Plants from dormant seed were allowed to produce flowers but were removed before they could produce seed to ensure they did not contaminate plots, so performance can only be assessed as survival to flower. In 2015 plants were removed as they emerged so cannot be included in performance analyses. All analyses include the subplot mean of the number of main-stem nodes at the first observation for each plant ('av.leafnodes') as a covariate to account for higher survival probabilities the later in life a plant is found. For each transect we used two models to address this question; in both cases the response was the proportion of emerged plants that flowered (propFlower). Models were reduced as in Q1, but only 'Yr.emergence' was retained if not significant.

Model 1) included all years and sites with plots of more than one age (i.e. excluding 2011 when all seeds were in their first summer, and HB-edge where all seeds are the same age). We used mixed effects binomial GLMMs (glmer command, lme4 package) with three random intercepts: plotID to account for plot effects, subplotID as subplots were resampled in multiple years and these observations are not independent (this also accounts for overdispersion in the data), and year to account for performance differences among years. To test whether time spent dormant affects performance differently among sites or sources, as one might expect if site-specific conditions deteriorate seeds differently or if some seed sources are better able to withstand time spent underground, initial models included interactions between Year of emergence and Site, and Year of emergence and Source. Significance of interactions and factors was determined using likelihood ratio tests.

Full model 1: $\text{propFlower} \sim \text{Yr.emergence} \times \text{Site} + \text{Yr.emergence} \times \text{Source} + \text{av.leafnodes} + (1|\text{year}) + (1|\text{plotID}) + (1|\text{subplotID})$

Model 2) used 2013 data only, as this year included both 1st year plants and 2 years of dormant plots, and dormant plots were checked frequently throughout the growing season. As HB-edge and NK-above1 had no first-year plants in 2013, they were excluded from this analysis. Binomial GLMMs included random intercepts for plotID and subplotID to account for plot effects and overdispersion.

Full model 2: $\text{propFlower} \sim \text{Yr.emergence} \times \text{Site} + \text{Yr.emergence} \times \text{Source} + \text{av.leafnodes} + (1|\text{plotID}) + (1|\text{subplotID})$

Nakiska

Performance (proportion of emerged plants to flower) was negatively affected by the number of years a seed lay dormant, but this effect did not differ among sites or sources.

Model 1 (NK edge, above1, above2; 2012-2014) MAM:

$\text{propFlower} \sim \text{Yr.emergence} + \text{Site} + \text{Source} + \text{av.leafnodes} + (1|\text{year}) + (1|\text{plotID}) + (1|\text{subplotID})$

Yr.emergence: -0.38 ± 0.13 , likelihood ratio $\chi^2_{df=1}=7.92$, $P = 0.0049$

Model 2 (NK edge, above2; 2013 only) MAM:

$\text{propFlower} \sim \text{Yr.emergence} + \text{Site} + \text{av.leafnodes} + (1|\text{year}) + (1|\text{plotID}) + (1|\text{subplotID})$

Yr.emergence: -0.42 ± 0.13 , likelihood ratio $\chi^2_{df=1}=9.61$, $P = 0.0019$

Hailstone Butte

In contrast to Nakiska, performance was not affected by the number of years seeds spent dormant, either across 2012-2014 or considering just 2013 (when monitoring of dormant plants was most frequent):

Model 1 (HB-high, HB-above; 2012-2014) MAM:

$\text{propFlower} \sim \text{Yr.emergence} + \text{Source} + \text{av.leafnodes} +$
 $(1|\text{year}) + (1|\text{plotID}) + (1|\text{subplotID})$

Yr.emergence: -0.28 ± 0.20 , likelihood ratio $\chi^2_{df=1}=1.99$, $P = 0.16$

Model 2 (HB-high, HB-above; 2013 only) MAM:

$\text{propFlower} \sim \text{Yr.emergence} + \text{av.leafnodes} + (1|\text{plotID}) + (1|\text{subplotID})$

Yr.emergence: -0.25 ± 0.18 , $\chi^2_{df=1}=1.94$, $P = 0.16$

1.7. Additional Performance Results

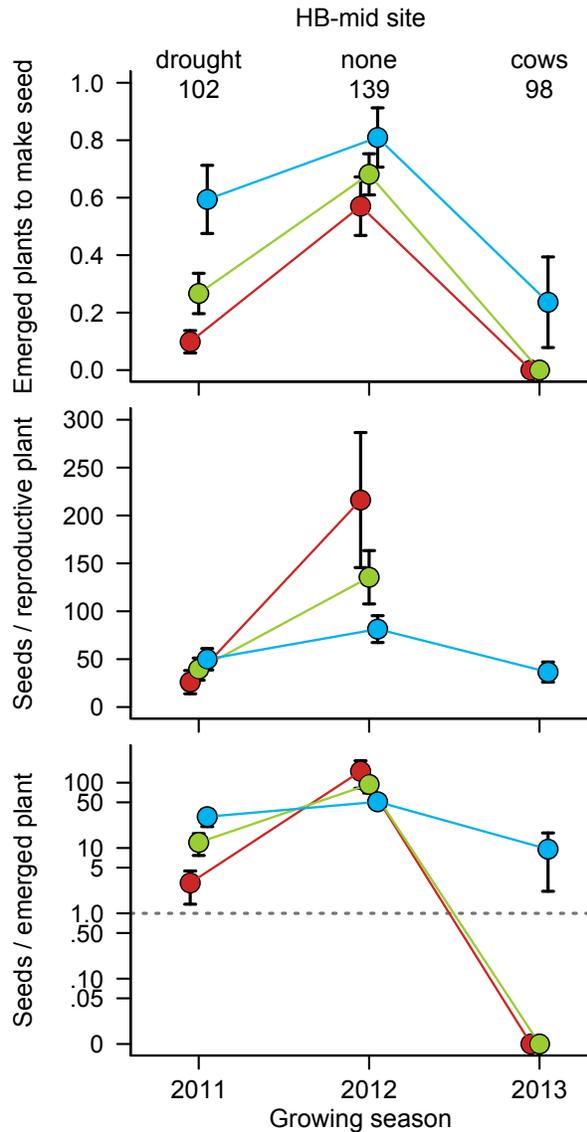


Fig. S6. Performance at HB-mid site in normal vs. truncated growing seasons (mean \pm SE). Text at top of panels indicates the type of disturbance and the number of days in the growing season from the start of *R. minor* emergence to the time at which >80% of the plants were killed or damaged. 2011: the fire tower at HB-above recorded 65 mm of rain in the 2011 growing season, compared to a mean of 240 mm (range 202–420 mm) in growing seasons from 2012–2015. Increasing drought meant that both transplant and natural *R. minor* plants died with buds and flowers still on them, with later plants the most severely affected. In 2013 cattle entered the site a month earlier than usual as fences were washed out in an extreme flood. In both cases the disturbance began after all (2011) or most (2013) high plants had set seed, and so affect high plants the least and low plants the most. Dashed line shows the minimum post-emergence success required for self-sustaining populations, assuming all seeds germinated. Lifetime reproductive success in the normal year (2012) is similar to that recorded for natural plants at the site during 2010 and 2015 (unpublished data).

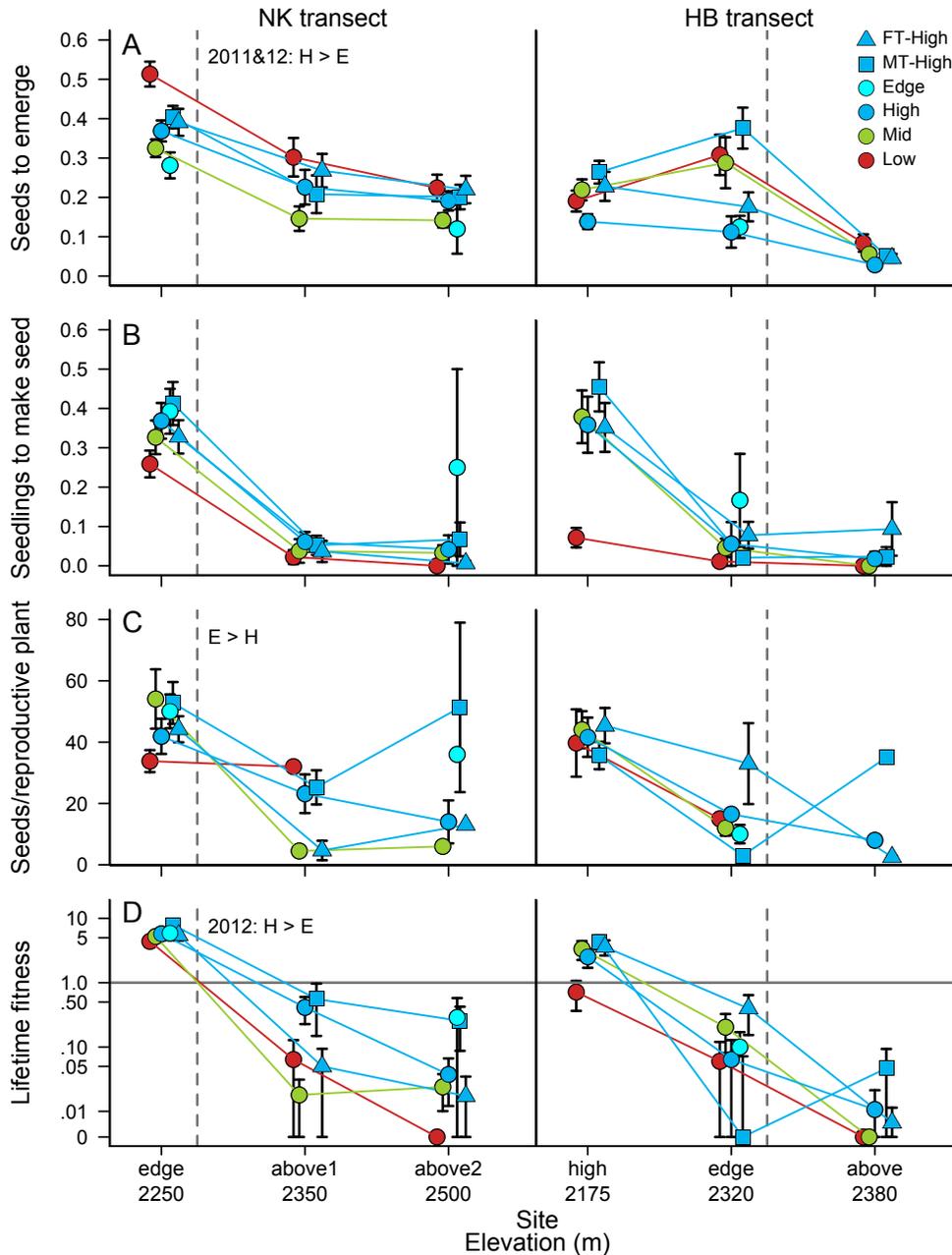


Fig. S7. Performance (mean \pm SE) of seeds from the absolute range edge at sites spanning *R. minor*'s high-elevation range limit (vertical dashed line). Models compare local edge seed performance to that of seeds from local high-elevation populations 100 m below the absolute range edge, which were used for reciprocal transplants; other sources are shown for reference. * indicates sites where high and edge sources differed in some years: 2011 and 2012 for emergence, 2012 for seed production (full statistical results in Table S5). All available data are shown for each source and site: all sources in all three years (2011–2013) at NK-edge, HB-high, HB-above sites; all sources in 2011–2012 only at NK-above1 and 2011 only at HB-edge; edge source in 2013 only but other sources in 2011–2013 at NK-above2.

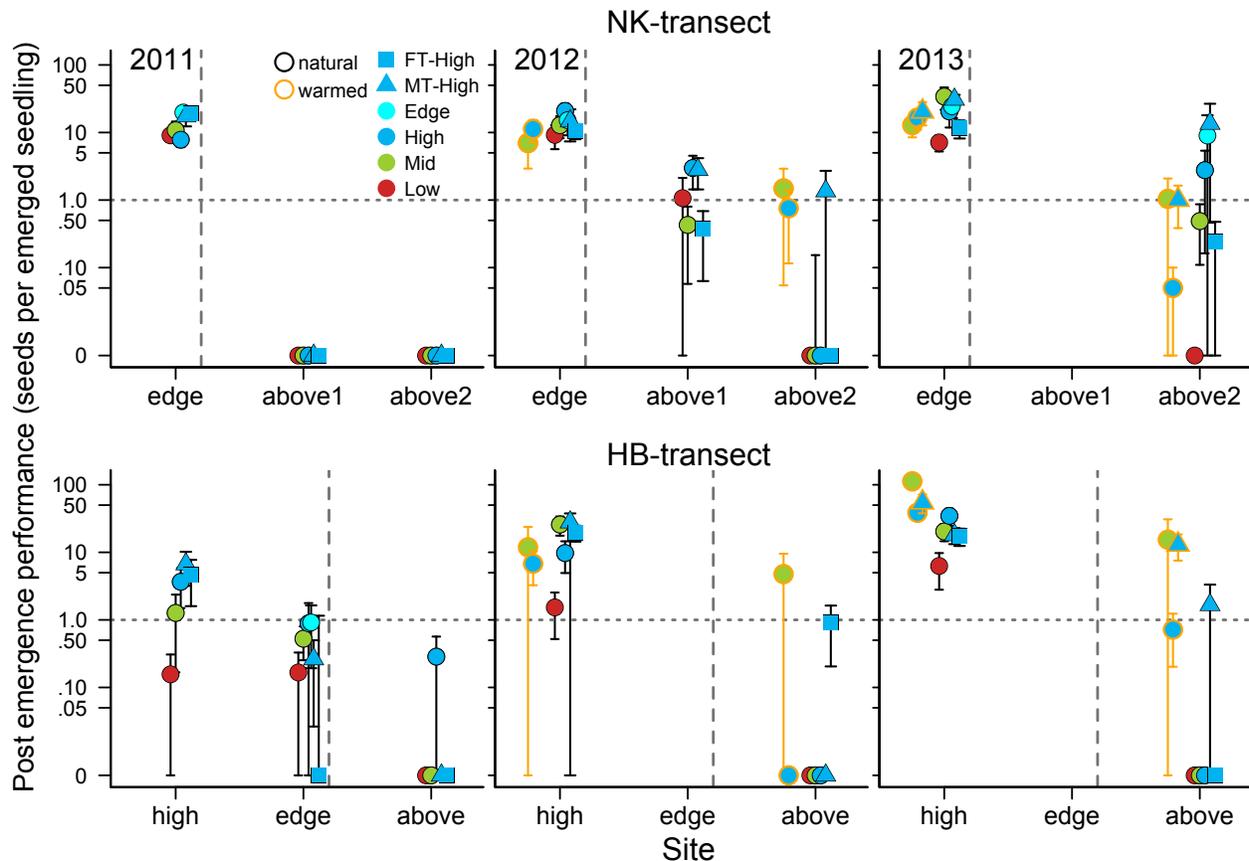


Fig. S8. Post-emergence performance by year across *R. minor*'s high-elevation range limit (vertical dashed line) under natural and artificially warmed conditions. Dashed reference line at 1 shows the minimum performance required for self-sustaining populations, assuming all seeds germinated. Data are first averaged by subplot, points show mean across subplots \pm SE. Because subplots can only be included if at least one seed emerged, sample sizes are sometimes less than the 10 subplots planted / source / treatment / year, especially at the highest sites where germination was low. Within the range (sites left of dashed line), n was almost always 9 or 10 subplots, except for warmed plots at HB-high in 2012, for which $n = 3$ as most plots planted for warming were retained in the control treatment due to low emergence. Above the range at NK, $n = 8-10$ for 2011-2012, but varied from 4 to 9 in 2013 (NK-above2). At HB-above emergence was much lower, ranging from 4 to 6 in 2011, 2 to 6 in 2012, and 5 to 12 in 2013 after extra plots were planted to make up for low emergence. Exact sample sizes are in Table S2b.

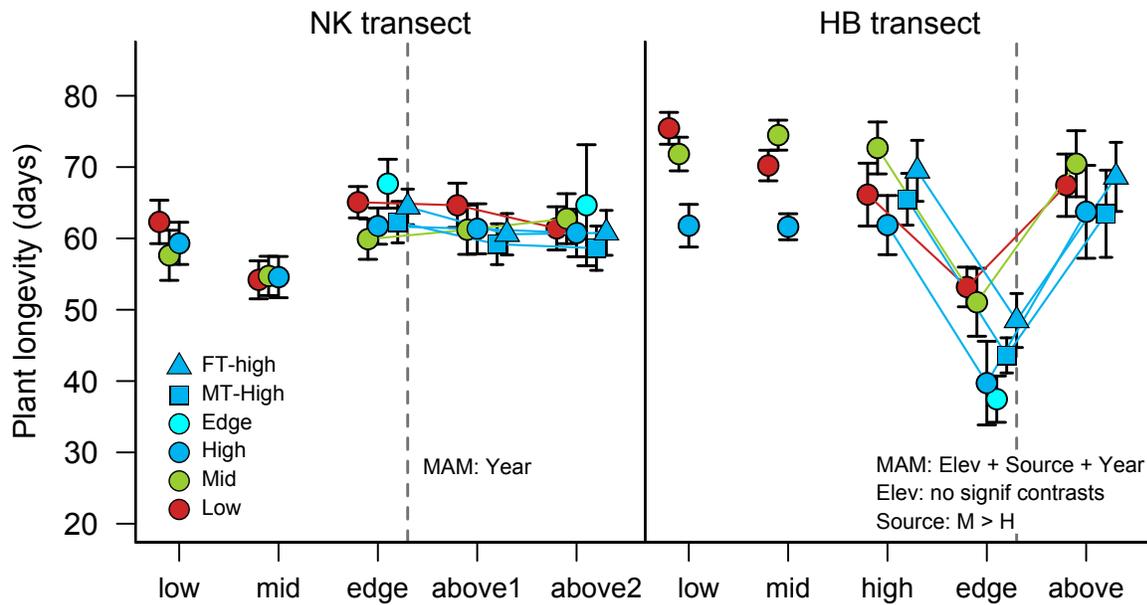


Fig. S9. Plant lifespan at and above the range edge (date of final fitness assessment – estimated date of emergence); mean \pm SE across years. We tested for differences in longevity among all sources at the highest three sites along each transect using a GLMM that left out the Site-Elevation \times Year interaction (full model: `longevity ~ Elevation x Source + Source x Year + (1|plot) + (1|subplot)`), data shown for low and mid sites for reference only). Significance of main effects and interactions was assessed using likelihood ratio tests. Non-significant interactions and main effects were dropped sequentially, and the adequacy of the minimum adequate model (MAM, given in text on figure) was confirmed using a likelihood-ratio test compared to the full model. When the MAM included Site Elevation or Source, differences among levels were assessed using least squared means (`lsmeans` function, `lsmeans` package).

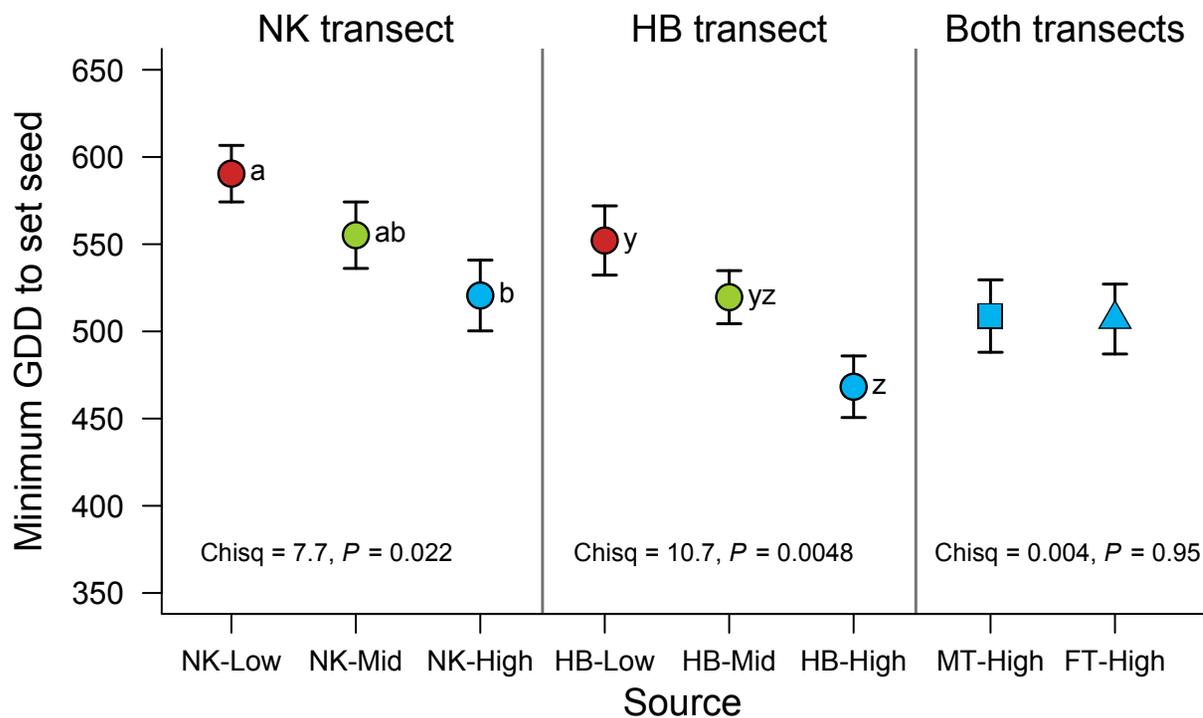


Fig. S10. Minimum growing degree days needed for each source to mature seed. Minimum GDD (mean \pm SE) is calculated for each source-site-year combination as the sum of the GDD between the date the first plant emerged and the date the first plant produced mature seed. Means are calculated across sites and years, including only sites within *R. minor*'s range as not all sources produced seed above the range. As sources were planted on different combinations of transects, source differences were analysed by transect (NK, HB, or both) using linear mixed models with source as a fixed effect and site as a random effect, fit with maximum likelihood (lmer function; lme4 package, 1.1-13 (Bates *et al.* 2015)). The importance of source was tested using a likelihood-ratio test comparing the full model to a null model with only the random effect and an intercept, using a χ^2 distribution. When source was significant we tested for pairwise differences among sources using the lsmeans and contrast functions (lsmeans package, 2.25-5); means not sharing a letter are significantly different from each other.

2. Statistical Results

Table S5. Analyses of performance (Fig. S7) and phenology of seeds from *R. minor*'s absolute range edge (E) vs. the highest large population (H) on two elevational transects (NK & HB). H seeds were collected 100 m downslope of E and were those used in reciprocal transplants. Transects are analysed separately using GLMMs with plot and subplot as random effects, and binomidal (proportional data) or Poisson (seed production, lifetime fitness, and phenology) error distributions. Models include only data from sites and years where edge seeds were planted: NK-edge (2250 masl) in 2011–2013, NK-above2 (2500 masl) in 2013, and HB-edge (2320 masl) in 2011. NK models have fixed effects Source, Site-Elevation, and Year (abbreviated to S, E, Y), and full models include all fully-crossed interactions. HB models have fixed effect Source. Term significance was assessed using likelihood ratio tests comparing models with and without a given term and a χ^2 distribution. Interactions and Year were dropped from models if not significant. Sample sizes (*n*) are the # subplots/source/site across all years considered in the model. Only differences among sources are shown here, as site effects are fully explored with the full data set (Tables S6-8).

Response		Site (<i>n</i> H, E)	Fixed effects		χ^2 test statistic, <i>df</i> , <i>P</i>			Significant Source contrasts
Years	Full model		Final model	Source	Source x Site-Elevation	Source x Year		
Proportion seeds to emerge								
NK	2011-2013 2013	E (60, 30) ¹ A2 (20, 5)	SxE + SxY	SxY + E	—	—	15.2, <i>df</i> =2 <i>P</i> = 0.0005	2011: H > E 2012: H > E
HB	2011	E (10, 10)	S	S	0.53, <i>df</i> =1 <i>P</i> = 0.47	—	—	—
Proportion seedlings to make seed								
NK	2011-2013 2013	E (30, 30) A2 (10, 5)	SxE + SxY	S + E	0.19, <i>df</i> =1 <i>P</i> = 0.66	—	—	—
HB	2011	E (10, 10)	S	S	0.39, <i>df</i> =1 <i>P</i> = 0.53	—	—	—
Seeds produced per reproductive plant (in-range sites only)²								
NK	2011-2013	E (30, 30)	SxE + SxY	S + E + Y	4.94, <i>df</i> =1 <i>P</i> = 0.026	—	—	E > H
HB	2011	E (10, 10)	S	S	3.18, <i>df</i> =1 <i>P</i> = 0.075	—	—	—
<i>continued...</i>								

Response	Years	Site (n H, E)	Fixed effects		χ^2 test statistic, df, P			Significant Source contrasts
			Full model	Final model	Source	Source x Site-Elevation	Source x year	
Lifetime fitness (seeds/seed planted)								
NK	2011-2013 2013	E (30, 30) A2 (10, 5)	SxE + SxY	SxY + E	—	—	10.5, df=2 P = 0.0053	2012: H > E
HB	2011	E (10, 10)	S ³	S	0, df=1 P = 1	—	—	—
Emergence date								
NK	2011-2013 2013	E (30, 30) A2 (10, 5)	ExS + SxY	E + S + Y	0.015, df=1 P = 0.90	—	—	—
HB	2011	E (10, 10)	S	S	0.036, df=1 P = 0.85	—	—	—
Start-of-flowering date								
NK	2011-2013 2013	E (30, 30) A2 (10, 5)	ExS + SxY	E + S + Y	0.059, df=1 P = 0.81	—	—	—
HB	2011	E (10, 10)	S	S	0.005, df=1 P = 0.95	—	—	—

- 1) *n* higher for emergence than other life stages as it includes plots later put in the warming treatment
- 2) Seeds/reproductive plant could not be analysed at NK-above2 as only one edge plant made seeds
- 3) Poisson models round response values to integers, so subplots that produce <0.5 seeds/seed planted are rounded to zero. Under this rounding regime H and E sources have identical performance (nine subplots with 0 seeds/seed planted and one with 1 seeds/seed planted). If we instead round any seed-producing subplot to LTF=1, there are eight E subplots with 0 seeds/seed planted and two with 1 seeds/seed planted, $\chi^2_{df=1} = 0.34$, $P = 0.56$.

Table S6. Phenology analyses (Fig. 4). Response variables are subplot means. Models consider local low, mid, and high-elevation (L, M, H) sources from the home transect transplanted to low (L), mid (M), high (H), range-edge (E), and above-range sites (A, A1, A2). Models were Poisson GLMMs. HB-E and NK-A1 sites were only planted for 2011 and 2011-2012, respectively, so full model fixed effects exclude interactions between Site-Elevation and Year: Site-Elevation \times Source + Source \times Year (ExS + SxY; separate models by year yield the same patterns but differences in start of flowering date between sources were not always significant due to lack of power). Term significance was assessed using likelihood ratio χ^2 tests comparing models with and without a given term. Non-significant terms were dropped from models, except for ‘Site-Elevation’ and ‘Source’, which were always retained in final models as they were integral to the experimental design. Estimated phenology dates take into account both the first date a plant was observed at a given life stage (emerged or flowering), and the quantitative phenological state of the plant (how many primary leaf nodes or primary flowers the plant had when first observed). E.g. a plant first seen with 1 primary leaf node would be given a later emergence date than a plant first seen on the same day but which had 6 leaf nodes (see main manuscript for estimate details). Analyses using only the observed date of first flowering (e.g. plants first seen flowering on the same day would have the same start-of-flowering date no matter how many flowers they had) yield the same significant contrasts.

Response		Fixed effects		χ^2 Test statistic, <i>df</i> , <i>P</i>				Significant contrasts	
		Full model	Final model	Site Elevation	Source	Year	Interact'n	Site Elevation	Source
Years	Sites								
Emergence date (estimated)									
NK 2011-13	L, M, E, A1, A2	ExS + SxY	E+S+Y	277.4, <i>df</i> =4 <i>P</i> < 0.001	0.04, <i>df</i> =2 <i>P</i> = 0.98	107.3, <i>df</i> =2 <i>P</i> < 0.001	—	L < M < E < A1=A2	—
HB 2011-13	L, M, H, E, A	ExS + SxY	E+S+Y	216.0, <i>df</i> =4 <i>P</i> < 0.001	0.50, <i>df</i> =2 <i>P</i> = 0.78	121.1, <i>df</i> =2 <i>P</i> < 0.001	—	La Mb Hc Ecd Ad	—
Start-of-flowering date (estimated)									
NK 2011-13	L, M, E, A1, A2	ExS + SxY	E+S+Y	222.2, <i>df</i> =4 <i>P</i> < 0.001	11.1, <i>df</i> =2 <i>P</i> = 0.004	18.5, <i>df</i> =2 <i>P</i> < 0.001	—	L < M < E < A1=A2	M=H < L
HB 2011-13	L, M, H, E, A	ExS + SxY	E+S+Y	216.4, <i>df</i> =4 <i>P</i> < 0.001	40.8, <i>df</i> =2 <i>P</i> < 0.001	8.8, <i>df</i> =2 <i>P</i> = 0.012	—	La Mb Hc Ecd Ad	H < M < L

Table S7. Analyses of performance across *R. minor*'s high-elevation range limit (Figs 2&3). Results from GLMMs with random effects for plot and subplot and binomial (proportional data) or Poisson (seeds/seed) error structure. Models consider local low, mid, and high-elevation (L, M, H) sources from that transect plus two high-elevation sources from nearby mountains (MT, F), transplanted to high (H), range-edge (E), and above-range sites (A, A1, A2). HB-E and NK-A1 sites were only planted for 2011 and 2011-2012, respectively, so full models exclude interactions between Site-Elevation and Year: Site-Elevation \times Source + Source \times Year (ExS + SxY; separate models for all years excluding these sites or all sites excluding some years yield the same significant contrasts). Term significance was assessed using likelihood ratio χ^2 tests comparing models with and without a given term. Non-significant terms were dropped from models. Sample sizes (*n*) are the maximum number of subplots/source at each site across all years considered in the model. Bold contrasts are consistent with local adaptation (i.e. at high-elevation sites high-elevation sources do best).

Response	Years	Sites <i>n</i> max	Final model	χ^2 test statistic, df, <i>P</i>				Significant contrasts	
				Site-Elev.	Source	Year	Interactions	Site-Elevation	Source
Proportion seeds to emerge									
NK	2011-13	E, A1, A2 50, 20, 50	E + S	46.0, df=2 <i>P</i> < 0.001	28.9, df=4 <i>P</i> < 0.001	—	—	E > A1=A2	L > H=MT=M
HB	2011-13	H, E, A 50, 10, 60	ExS + SxY	—	—	—	ExS: 20.4, df=8 <i>P</i> = 0.0088 SxY: 32.0, df=8 <i>P</i> < 0.001	forL,M,MT,F: E=H > A forH: H > A	2011: L > H 2012: La MTa Fab Mb Hc 2013: F=MT=M=L > H
Proportion seedlings to make seed¹									
NK	2011-13	E, A1, A2 30, 20, 30	E + SxY	101.9, df=1 <i>P</i> < 0.001	—	—	SxY: 27.2, df=16 <i>P</i> = 0.039	E > A1=A2	2012: H > L 2013: MT > L
HB	2011-13	H, E, A 37, 10, 39	E + S + Y	42.7, df=2 <i>P</i> < 0.001	57.3, df=4 <i>P</i> < 0.001	10.1, df=2 <i>P</i> =0.0064	—	H > E=A	H=MT=F=M > L
Lifetime fitness (seeds/seed planted)									
NK	2011-13	E, A1, A2 30, 20, 30	ExS + SxY	—	—	—	ExS: 21.7, df=8 <i>P</i> = 0.0054 SxY: 25.3, df=8 <i>P</i> = 0.0014	E > A1=A2	2011: MT=F=H > L=M 2012: H=MT=F > L=M 2013: MT > H=F > L=M
HB	2011-13	H, E, A 30, 10, 30	E + S + Y	68.6, df=2 <i>P</i> < 0.001	32.3, df=4 <i>P</i> < 0.001	15.1, df=2 <i>P</i> < 0.001	—	H=E > A	MT=F=M=H > L

1) Not all sources have the maximum *n* for this response due to zero emergence in some above-range subplots– see Table S2b.

Table S8. Analyses of performance within the range (Figs 2&3). Results from binomial (b) or Poisson (p) GLMMs, comparing local low, mid, and high-elevation (L, M, H) sources at low, mid, high and range-edge transplant sites (L, M, H, E); all models include a random intercept for plot to account for plot-level effects, and for subplot to account for over or under dispersion. Full model fixed effects are Site-Elevation \times Source \times Year (ExSxY) for NK models (where Site-Elevation and Year are fully crossed), and ExS + SxY for HB models (Site-Elevation and Year are not fully crossed as the HB-edge site was only planted for 2011, so ExY interactions are excluded). Separate models for 2011-13 for L,M,H sites and 2011 for H and E sites yield the same significant contrasts). Sample sizes (n) are the maximum number of subplots/source at each site across all years considered in the model (n given by source in Table S2). Model reduction, sample sizes and contrasts are as for Table S5 – for design details see Table S1. Contrasts indicative of local adaptation (i.e. home site advantage) are in bold; contrasts always list factor levels from largest to smallest.

Response (error distribution)			χ^2 test statistic, df, P					Significant contrasts	
Years	Sites max n /Source	Final model	Site Elevation	Source	Year	Interactions	Site Elevation	Source	
Proportion seeds to emerge (b)¹									
NK	2011-13	L, M, E 25, 35, 60	ExY + S	—	52.0, df=2 $P < 0.001$	—	43.1, df=4 $P < 0.001$	2011: E=M > L 2012: M=E > L 2013: M > E=L	L > H=M
HB	2011-13	L, M, H, E 25, 34, 50, 10	E + SxY	22.1, df=3 $P < 0.001$	—	—	SxY: 22.8, df=4 $P < 0.001$	M=L > H=E ²	2011: M=L > H 2012: L > M > H 2013: L=M > H
Proportion seedlings to make seed (b)¹									
NK	2011-13	L, M, E 25, 25, 30	ExY + S	—	11.7, df=2 $P = 0.0028$	—	16.6, df=4 $P = 0.0024$	2013: L=E > M	H > L
HB	2011-13	L, M, H, E 25, 34, 37, 10	ExS + Y	—	—	7.9, df=2 $P = 0.02$	ExS: 22.7, df=6 $P < 0.001$	forL: La Mab Hb Ec forM: L=H=M > E forH: Ma Lab Hb Ec	atM: H > M > L atH: H=M > L
Seeds produced per reproductive plant (p)									
NK	2011-13	L, M, E 25, 25, 30	ExY + SxY	—	—	—	ExY: 11.5, df=4 $P = 0.022$ SxY: 10.0, df=4 $P = 0.040$	2011: L > M=E	2013: M > L

continued...

HB	2011-13	L, M, H, E 24, 30, 27, 10	ExS + SxY	—	—	—	ExS: 14.9, $df=6$ $P = 0.021$ SxY: 12.4, $df=4$ $P = 0.015$	forL: Ma La Eab Hb forH: M > L	atL: L=M > H
Lifetime fitness (seeds/seed planted) (p)									
NK	2011-13	L, M, E 25, 25, 30	ExSxY	—	—	—	ExSxY: 22.7, $df=8$ $P = 0.0037$	2012 forL: M=E > L forM: E=M > L forH: E > M=L 2013 forL: L > M forM: E > M	2011: atL: M > L
HB	2011-13	L, M, H, E 25, 34, 30, 10	ExS + SxY	—	—	—	ExS: 20.3, $df=6$ $P = 0.0024$ SxY: 41.7, $df=4$ $P < 0.001$	forL: L=M > H=E forM: Ma La Hab Eb forH: Ma Lb Hbc Ec	atL: L=M > H atH: M=H > L

-
- 1) In 2011 sites were at different phenological stages when first checked (plants were older at low and mid vs. higher sites). First checking a site at a relatively later phenology stage misses very early mortality, biasing emergence estimates downward and survival estimates upward. To standardize among sites, we determined the phenological stage at which the lowest (i.e. most advanced) site was first checked, and the comparable date at each higher site when plants were of a similar age (the equivalent-first-check date). We calculated ‘standardized emergence’ by excluding plants that failed before the equivalent-first-check date, and used this to calculate standardized proportional emergence and proportion of emerged seedlings to survive to make seed.
 - 2) The difference in emergence between Low and Edge sites is borderline significant, $P=0.052$

Table S9. Analyses of performance (Fig. 5) and phenology for the warming experiment. Experimental warming was done using open-topped chambers (OTCs) in 2012-2013 at a site-elevation close to the range edge (nearE: NK-edge and HB-high) and above the range (A: NK-above2 and HB-above) on each transect. In 2012 we warmed source populations from the mid (M) and high-elevation (H) populations local to each transect, and in 2013 we added a high-elevation source from nearby Moose Mt (MT-High). Analyses consider Warming treatment, Site Elevation, Source, Transect, and Year as fixed factors (abbreviated to W, E, S, T, Y). We model Year as a fixed effect because it has too few levels to be well-modeled as a random effect, but since Year is not fully crossed with Source and we are only interested in its effect on the warming treatment, we only model its interaction with the warming treatment. Thus the full model fixed effects are $OTC \times E \times T \times S + OTC \times Y$, except for ‘Seeds/plant’ where the 4-way interaction was excluded as no unwarmed M or H plants made seed at the HB-A site. Modeling year as a random intercept yielded the same final models and the same significant contrasts between OTC and CT treatments. As in other analyses, GLMMs considered subplot means as the response variable, a random intercept for subplot to account for overdispersion, and binomial or Poisson error distributions for proportion vs seed-count data. When planting for this experiment we paired each control plot with an OTC plot, such that both plots experienced similar vegetation and terrain (except the 10 extra plots planted at HB-above in 2013, which were unpaired). We therefore include plot-pairID as an additional random intercept. Sample sizes (# subplots per factor level) are given for each factor in each analysis in brackets at the bottom of the cell. Results for ‘Seeds per emerged plant’ (i.e. the performance parameter closest to lifetime fitness) are shown in Fig. 5.

Response (error dist.)	χ^2 test statistic, $df=1$, P					Significant contrasts between OTC vs CT treatments
	(n subplots per factor level)					
Final model fixed effects (<i>n</i> total subplots)	Warming (OTC, CT)	Elevation (nearE, A)	Source $df=2$ (M, H, MT)	Transect (NK, HB)	Warming Interactions	
Proportion emerged plants to flower (b)						
WxE + T (<i>n</i> =329)	— (155, 174)	— (185, 144)	2.7, $P = 0.25$ (133, 121, 75)	8.4, $P = 0.0038$ (178, 151)	W x E: 6.5, $P = 0.011$	atA: OTC > CT
Proportion flowering plants to make seed (b)						
WxT + WxE + S (<i>n</i> =248)	— (123, 125)	— (164, 84)	8.9, $P = 0.012$ (102, 85, 61)	— (137, 111)	W x T: 27.2 $P < 0.001$ W x E: 3.9, $P = 0.047$	atHB: OTC > CT atH: OTC > CT atA: OTC > CT
Seeds per seed-making plant (p)						
WxSxT + WxE x S + Y (<i>n</i> =181)	— (91, 90)	— (153, 28)	— (67, 65, 49)	— (102, 79)	W x S x T: 11.8, $P = 0.0028$ W x E x S: 11.1, $P = 0.0040$	atNK forMT: CT > OTC atHB forM: OTC > CT atA forM: OTC > CT
Start-of-flowering date (estimated*) (p)						
W + E + S + T (<i>n</i> =244)	1.5, $P = 0.22$ (121, 123)	107.1, $P < 0.0001$ (162, 82)	3.5, $P = 0.17$ (102, 82, 60)	13.5, $P < 0.001$ (136, 108)	—	—
Seeds per emerged plant (LTF equivalent) (p)						
WxE + WxT + ExS + SxT + Y (<i>n</i> =329)	— (155, 174)	— (185, 144)	— (133, 121, 75)	— (178, 151)	W x E: 5.4, $P = 0.020$ W x T: 25.3, $P < 0.001$	atA: OTC > CT atHB: OTC > CT

1) The estimated date of first flower takes into account both the date the plant was first observed with primary-stem flowers or finished flowers and the number of primary flowers the plant had on this date. Using the observed date (i.e. a coarser comparison) yields the same final model and lack of warming-treatment effect.

3. Additional Discussion

Interpreting the effects of OTCs. Open-topped warming chambers (OTCs) directly test whether insufficient heat accumulation (measured as GDD) during the growing season constrains post-emergence fitness across *R. minor*'s upper range edge, but only partially mimic the effects of climate change. Whereas OTCs primarily warmed daytime temperatures, climate change is primarily warming nighttime temperatures (Davy *et al.* 2017). Climate change will also affect winter conditions, likely advancing spring melt and altering snowpack depth (Barnett *et al.* 2005). Earlier melt would add GDD to growing seasons. Extra GDD via longer growing seasons would presumably have the same positive effect on fitness above the range as increasing GDD via warming. In contrast, deeper/shallower snow could improve/worsen fitness if poor snow coverage caused low emergence at ridge-top sites (Figs S2B & 2A). While all of these additional effects could modify the effect of warmer growing season temperatures (the primary effect of OTCs) on plant fitness, none seem likely to restore the advantage of high elevation genotypes. How climate change will alter soil moisture in our study area remains unclear (Barnett *et al.* 2005; CCCMA 2017), and OTC effects on soil moisture were unmeasured, so their correspondence is unknown. Finally, *R. minor* will not migrate upslope alone, and vegetation at our above-range sites may increasingly resemble that in *R. minor*'s range. If anything, we expect this to act like warming, improving fitness for all genotypes but reducing the advantage of high vs. mid-elevation seeds (Fig. 3). So, while OTCs imperfectly mimic climate-change, we expect their main predictions—that warming improves fitness but eliminates the advantage of high-elevation over mid-elevation seeds—to hold.

References

- Barnett, T.P., Adam, J.C. & Lettenmaier, D.P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, 438, 303.
- Bates, D., Maechler, M., Bolker, B.M. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.*, 67, 1-48.
- CCCMA (2017). Canadian Centre for Climate Modelling and Analysis Climate forecasts: Canada. Environment and Climate Change Canada.
- Cesaraccio, C., Spano, D., Duce, P. & Snyder, R.L. (2001). An improved model for determining degree-day values from daily temperature data. *International journal of biometeorology*, 45, 161-169.
- Davy, R., Esau, I., Chernokulsky, A., Outten, S. & Zilitinkevich, S. (2017). Diurnal asymmetry to the observed global warming. *Int. J. Climatol.*, 37, 79-93.
- Donohue, K., Rubio de Casas, R., Burghardt, L., Kovach, K. & Willis, C.G. (2010). Germination, postgermination adaptation, and species ecological ranges. *Annu. Rev. Ecol. Evol. Syst.*, 41, 293-319.
- Hargreaves, A.L., Weiner, J.L. & Eckert, C.G. (2015). High-elevation range limit of an annual herb is neither caused nor reinforced by declining pollinator service. *J. Ecol.*, 103, 572-584.
- Lenth, R.V. (2016). Least-Squares Means: The R Package lsmeans. *J. Stat. Softw.*, 69, 1-33.
- Marion, G.M., Henry, G.H.R., Freckman, D.W., Johnstone, J., Jones, G., Jones, M.H. *et al.* (1997). Open-top designs for manipulating field temperature in high-latitude ecosystems. *Glob. Change Biol.*, 3, 20-32.
- R Development Core Team (2015). R: a language and environment for statistical computing. Westbury, D.B. (2004). *Rhinanthus minor* L. *J. Ecol.*, 92, 906-927.
- Womach, J. (2005). Agriculture: a glossary of terms, programs, and laws, 2005 Edition. (ed. Service, CR). Library of Congress Washington.
- Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A. & Smith, G.M. (2009). *Mixed effects models and extensions in ecology with R*. Springer, New York, NY.